

Postdoc

Multi-objective Surrogate Performance Prediction Models for Machine Learning Applications in Hardware Design Space Exploration

This Post-doc will be done in the ConfianceAI framework ([Confiance.ai](https://confiance.ai)). Confiance.ai is the technological pillar of the Grand Défi "Securing, certifying and enhancing the reliability of systems based on artificial intelligence" launched by the French Innovation Council. It is the largest technological research programme in the #AIforHumanity plan, which is designed to make France one of the leading countries in artificial intelligence (AI).

This post-doc aims to design models to estimate and predict CNN performances without requiring their deployment on the real platform. These models will save time since measurement campaigns will be needed. Furthermore, it will be possible to test CNN architectures and measure their performances, in terms of execution time, power consumption, memory occupation, etc, before the platforms being available in the market.

Thus, measurements such as execution time, or energy consumption can be estimated on an architecture not available from the manufacturer if we know its most "impacting" attributes on the Hardware architecture(s).

This solution makes it possible to quickly explore a large space of configurations (CNN architectures, hardware architectures). It also makes it possible to avoid developing and optimizing models that will turn out ineffective on the target platform. In this postdoc we will develop new models for CNN performance prediction that respect the rank of the models (Pareto rank-preserving surrogate models).

Hardware-aware Neural Architecture Search (HW-NAS) has recently gained steam by automating the design of efficient DL models. However, such algorithms require excessive computational resources and thousands of GPU days are needed to evaluate and explore CNN search space. In this Postdoc we will explore the effectiveness of multi-objective Surrogate Performance Prediction Models in HW-NAS.

Tasks that have to be developed in this postdoc are :

- Determination of the CNN search space
- Determination of the main characteristics for two of Hardware platforms. This may concern Qualcomm Hexagon DSPs and edge Nvidia GPUs.
- Development and testing of multi-objective surrogate performance prediction models for these 2 types of HW platforms.

Education: A Ph.D in computer/electrical science/engineering is required

Salary: 2300 euros/month net (3000 euros/month gross)

Deadline for application: 30/09/2022

Duration : 1 year (1 year extension possible)

Employer : IRT Systemx

Address: 2 Bd Thomas Gobert, 91120 Palaiseau

Start date: Preferably between Sept 1st 2022 and Nov 1th 2022.

An application prepared in English or French should contain:

1. CV with the list of publications.
2. Contact information for 2-3 reference persons.
3. Your most relevant conference or journal publications, in full-text.

For further information contact:

- Prof. Smail Niar (smail.niar@uphf.fr) LAMIH/CNRS, INSA Hauts-de-France and CNRS, France
- Prof Elghazali Talbi (el-ghazali.talbi@univ-lille.fr), INRIA and University of Lille, France